An Information-Theoretic Derivation of the MMSE Decision-Feedback Equalizer

Tommy Guess and Mahesh K. Varanasi
Department of Electrical and Computer Engineering
University of Colorado, Boulder, CO 80309-0425

****

Thirty-Sixth Annual Allerton Conference on Communication, Control, and Computing
September 23-25, 1998

Abstract

We consider the discrete-time Gaussian channel with inter-symbol interference (ISI). Under the assumption of perfect feedback, an information-theoretic derivation of the minimum mean-squared error (MMSE) decision-feedback equalizer (DFE) is presented. Whereas previous works have used information theory to analyze the zero-forcing and MMSE DFE structures, this work *derives* the MMSE DFE directly by means of information-lossless projections applied to the average mutual information of the Gaussian ISI channel. Specifically, the perfect-feedback MMSE DFE works by performing an information-lossless conversion of the ISI channel, which must be viewed in a sequence-wise manner, into a sequence of input-output pairs that must be viewed symbol-wise. With ideal interleaving, the latter can also be viewed as a memoryless Gaussian channel. This equivalence resolves the paradoxical result that perfect post-cursor ISI cancellation is in general an information-increasing operation. We find that it is not perfect cancellation that increases the mutual information, but rather the mathematically inconsistent assumption that the perfect-feedback MMSE DFE can be viewed as a sequence-wise channel with memory.

I. INTRODUCTION

Consider the discrete-time Gaussian inter-symbol interference (ISI) channel where the output of the channel at the k^{th} time instance is given as

$$y_k = \sum_{j=-\infty}^{\infty} a_j x_{k-j} + n_k. \tag{1}$$

The input sequence $\{x_k\}$ and the noise sequence $\{n_k\}$ are both complex-valued, zero-mean wide-sense stationary (WSS) processes that are statistically independent of each other. Moreover, the noise sequence is assumed to be circularly symmetric[1] and Gaussian. The sequence $\{a_k\}$ is assumed to be in ℓ_2, i. e., $\sum_{k=-\infty}^{\infty} |a_k|^2 < \infty$, so the the output sequence $\{y_k\}$ is also

This work was supported in part by NSF grants NCR-9406069 and NCR-9706591.

[1] A random variable is circularly symmetric or proper if its *pseudo* variance, i. e., $E[(x - E[x])^2]$, as opposed to $E[|x - E[x]|^2]$, is equal to zero. Similarly a discrete-time random process is circularly symmetric or proper if $E[x_{k+l}x_k] = 0$ for all k and l. See [1, Sec. 8.1.1] and [2]. We will assume throughout this paper that all Gaussian variables and processes are complex-valued and circularly symmetric.

WSS. This channel arises naturally in the context of quadrature amplitude modulation (QAM) signaling over a time-dispersive Gaussian channel when the received signal is processed by a matched filter and sampled at the symbol rate (see, for example, [1, Sec. 6.2.1] [3, Sec. 6.2] [4]).

To facilitate the presentation, we will formally use the z-transform to represent a convolution operator. Thus for any sequence $\{h_k\}$, stochastic or deterministic, we write that

$$h(z) = \sum_{k=-\infty}^{\infty} h_k z^{-k}. \tag{2}$$

This allows the ISI channel in (1) to be expressed as

$$y(z) = a(z)x(z) + n(z). \tag{3}$$

Of course, $x(z)$, $n(z)$, and $y(z)$ do not exist as true z-transforms, but since $\{x_k\}$, $\{n_k\}$, and $\{y_k\}$ are WSS processes, their auto- and cross-correlation functions do possess z-transforms. For example, if we let $R_{xy}(k) = E[x_{l+k}y_l^*]$, then the cross spectrum of $x(z)$ and $y(z)$ is given by

$$S_{xy}(z) = \sum_{k=-\infty}^{\infty} R_{xy}(k) z^{-k}. \tag{4}$$

In fact, it is easy to see that

$$\begin{aligned} S_y(z) &= a(z)S_x(z)a^*(1/z^*) + S_n(z) && (5)\\ S_{yx}(z) &= a(z)S_x(z), && (6) \end{aligned}$$

since $x(z)$ and $n(z)$ are independent. We have used the notation $a^*(1/z^*)$ to represent $\sum_{k=-\infty}^{\infty} a_k^* z^k$.

The average mutual information of the ISI channel, which we denote by $I(x(z); y(z))$, is given by [5] [6] [7]

$$I(x(z); y(z)) = \lim_{N\to\infty} \frac{1}{N} I(x_1^N; y_1^N), \tag{7}$$

where we have used the notation

$$I(x_1^N; y_1^N) = I(x_1, x_2, \ldots, x_N; y_1, y_2, \ldots, y_N) \tag{8}$$

with the function $I(\cdot\,;\cdot)$ denoting mutual information.

If we are given that $E[|x_k|^2] = p$, then it is well known that the particular input process $\{x_k\}$ that maximizes the average mutual information of the Gaussian ISI channel will itself be a Gaussian process [5]. The spectrum $S_x(z)$ that maximizes the average mutual information is a function of both the channel $a(z)$ and the noise spectrum $S_n(z)$, and it is determined by means of the so-called "water filling" procedure. Thus, we will assume hereafter that $\{x_k\}$ is a Gaussian process. This being the case, it is known that

$$I(x(z); y(z)) = \int_{-\pi}^{\pi} \log\left(\frac{S_y(e^{j\theta})}{S_n(e^{j\theta})}\right) \frac{d\theta}{2\pi} \tag{9}$$

in nats per two dimensions, since the symbols are complex valued.

In the sections that follow, Section II reviews some important results from both the information theory of Gaussian processes and prediction theory. These properties are incorporated in Section III to derive the MMSE DFE directly from $I(x(z); y(z))$. This approach is in contrast to previous works that have started with various DFE structures and then analyzed them in an information-theoretic context [8] [1, Sec. 10.5.5] [9] [10] [11]. There are distinct advantages to the approach contained herein. First, it gives insight as to how the MMSE DFE functions. The MMSE DFE receiver essentially converts the ISI channel into a sequence of input-output pairs that must be viewed symbol wise; with ideal interleaving, this can be viewed as a memoryless Gaussian channel. This conversion is achieved by means of information-lossless projections that yield a set of sufficient statistics. Additionally, we are able to resolve the paradoxical idea that perfect feedback in a MMSE DFE receiver structure can actually lead to a “fictitious” increase in mutual information. This idea is reported and discussed in [11], [8], and [10]. For perfect post-cursor ISI cancellation to be information increasing, the effective channel yielded by the MMSE DFE receiver must be viewed as a sequence-wise channel with memory, something our derivation shows is mathematically inconsistent. This resolution allows us to state a new converse to the coding theorem for the perfect-feedback MMSE DFE receiver.

II. SOME KEY RESULTS FROM INFORMATION THEORY AND LINEAR PREDICTION

In this section we present several results from information theory and prediction theory that are used to derive the MMSE DFE in Section III.

A. The Mutual Information Between Gaussian Processes

Here we are concerned with the mutual information between Gaussian random variables and processes. Some of the results we review are well known, but others are not.

Lemma 1: Let $X = \{x_k\}_{k\in\mathcal{G}}$, $Y = \{y_k\}_{k\in\mathcal{H}}$, $Z = \{z_k\}_{k\in\mathcal{J}}$, and (X, Y, Z) each be a set of jointly Gaussian zero-mean random variables with the sets $\mathcal{G}$, $\mathcal{H}$, and $\mathcal{J}$ being arbitrary. Then the conditional mutual information

$$I(X; Y|Z) = I((\mathbf{E} - \mathbf{\Pi}_Z)(X); (\mathbf{E} - \mathbf{\Pi}_Z)(Y)), \tag{10}$$

where $\mathbf{E}$ denotes the identity operator and $\mathbf{\Pi}_Z$ denotes the operator that orthogonally projects on to Z. $\diamond$

This lemma is shown in [12, Sec. 9.3]. Since the operator $(\mathbf{E} - \mathbf{\Pi}_Z)$ yields that part of the argument which is orthogonal to Z, we see that the conditional mutual information has been expressed as the mutual information between the parts of X and Y that are orthogonal to Z.

Lemma 2: With the same hypotheses as Lemma 1, let $\mathbf{\Pi}_Y$ be the orthogonal-projection operator onto Y. Then we have that

$$I(X;Y) = I(X;\mathbf{\Pi}_Y(X)). \tag{11}$$

◇

The previous two lemmas follow naturally from the more fundamental result for Gaussian variables that the mutual information between X and Y is a function of only the angles between their associated Hilbert spaces [12, Sec. 9.2] [13]. For example, in Lemma 2 the non-zero angles between X and Y are the same as those between X and $\mathbf{\Pi}_Y(X)$ so that the mutual information is unaltered. And in Lemma 1, those parts of X and Y that lie in Z contribute nothing to the mutual information. The inner product of the Hilbert space is expectation, so for zero-mean Gaussian random variables x and y we have that

$$\mathbf{\Pi}_y(x) = E[x|y] = \frac{E[xy^*]}{E[|y|^2]}y. \tag{12}$$

The following lemma is a variation of the well known $\log(1+\mathrm{SNR})$ result for the mutual information between two zero-mean Gaussian random variables, i. e., the capacity of memoryless discrete-time Gaussian channel; it evaluates the signal-plus-noise power divided by the noise power:

Lemma 3: Given zero-mean jointly Gaussian random variables x and y,

$$I(x;y) = \log\left(\frac{E[|y|^2]}{E[|y-E[y|x]|^2]}\right). \tag{13}$$

◇

Lemma 4: Suppose that at least one of the zero-mean WSS processes $\{x_k\}$ and $\{y_k\}$ is regular (i. e., non-predictable). That is, $\int_{-\pi}^{\pi} |\log(S_x(e^{j\theta}))|d\theta < \infty$ or $\int_{-\pi}^{\pi} |\log(S_y(e^{j\theta}))|d\theta < \infty$. Then for all k we have that

$$I(x(z);y(z)) = I(x_k; y_{-\infty}^{\infty}|x_{-\infty}^{k-1}). \tag{14}$$

◇

For a discussion of this result see [12, Thm. 10.2.1].

B. *Results from Linear Prediction Theory*

Suppose we are given the regular (cf. Lemma 4), zero-mean WSS process $\{x_k\}$. The innovations process for $\{x_k\}$ is given by

$$(i_x)_k = x_k - E[x_k|x_{-\infty}^{k-1}]. \tag{15}$$

From linear prediction theory for scalar processes, we know that

$$x_k = \sum_{j=0}^{\infty}(\phi_x)_j(i_x)_{k-j}, \tag{16}$$

where $\phi_x(z) = \sum_{k=0}^{\infty}(\phi_x)_k z^{-k}$ is causal, monic (i. e., $(\phi_x)_0 = 1$), and minimum-phase [14, Secs.10-5, 13-4]. This means that $(\phi_x(z))^{-1}$, which we shall write as $\phi_x^{-1}(z)$, is also causal and monic. Moreover, we have that

$$S_x(z) = g_x \phi_x(z)\phi_x^*(1/z^*), \tag{17}$$

where $g_x = E[|(i_x)_k|^2]$ is the geometric mean of the spectrum $S_x(e^{j\theta})$. That is,

$$g_x = \exp\left(\int_{-\pi}^{\pi} \log(S_x(e^{j\theta}))\frac{d\theta}{2\pi}\right). \tag{18}$$

This allows us to formally write that $x(z) = \phi_x(z)i_x(z)$, and if we consider the formal z-transform of the sequence $\{\hat{x}_k\} = \{E[x_k|x_{-\infty}^{k-1}]\}$, we get that $\hat{x}(z) = x(z) - i_x(z) = (1 - \phi_x^{-1}(z))x(z)$. These results are summarized in the following lemma.

Lemma 5: Given a regular, zero-mean WSS process $\{x_k\}$, define the processes $\{\hat{x}_k\}$ and $\{(i_x)_k\}$ by $\hat{x}_k = E[x_k|x_{-\infty}^{k-1}]$ and $(i_x)_k = x_k - \hat{x}_k$. Then

$$\begin{aligned} x(z) &= \phi_x(z)i_x(z) && (19)\\ i_x(z) &= \phi_x^{-1}(z)x(z) && (20)\\ \hat{x}(z) &= (1 - \phi_x^{-1}(z))x(z). && (21)\end{aligned}$$

The polynomial $\phi_x(z)$ is causal, monic, and minimum phase; the spectral factorization of $S_x(z)/g_x$ is $\phi^*(1/z^*)\phi(z)$, where g_x given by (18). ◇

Having established the necessary background we proceed in the next section to derive the MMSE DFE receiver structure.

III. AN INFORMATION-THEORETIC DERIVATION OF THE MMSE DFE

We begin by proving the average mutual-information result previously stated in equation (9). Our approach parallels that given in [12, Sec. 10.2]. For the Gaussian ISI channel in (3) with a regular, zero-mean WSS Gaussian input, we consider $I(x(z); y(z))$. We know from Lemma 4 that

$$I(x(z); y(z)) = I(x_k; y_{-\infty}^{\infty}|x_{-\infty}^{k-1}) \tag{22}$$

for all k. Using the chain rule of mutual information, we can express $I(x_k; y_{-\infty}^{\infty}, x_{-\infty}^{k-1})$ in two different ways.

$$\begin{aligned} I(x_k; y_{-\infty}^{\infty}, x_{-\infty}^{k-1}) &= I(x_k; y_{-\infty}^{\infty}) + I(x_k; x_{-\infty}^{k-1}|y_{-\infty}^{\infty}) && (23)\\ I(x_k; y_{-\infty}^{\infty}, x_{-\infty}^{k-1}) &= I(x_k; x_{-\infty}^{k-1}) + I(x_k; y_{-\infty}^{\infty}|x_{-\infty}^{k-1}). && (24)\end{aligned}$$

Combining these two results with (22) allows us to see that

$$I(x(z); y(z)) = I(x_k; y_{-\infty}^{\infty}) + I(x_k; x_{-\infty}^{k-1}|y_{-\infty}^{\infty}) - I(x_k; x_{-\infty}^{k-1}). \tag{25}$$

Applying Lemma 2 to the first term and Lemma 1 to the second term we get

$$I(x(z);y(z)) = I(x_k; E[x_k|y_{-\infty}^{\infty}]) + I(x_k - E[x_k|y_{-\infty}^{\infty}]; \{x_j - E[x_j|y_{-\infty}^{\infty}]\}_{j=-\infty}^{k-1}) - I(x_k; x_{-\infty}^{k-1}). \quad (26)$$

Define the sequences $\{\tilde{x}_k\}$ and $\{e_k\}$ by $\tilde{x}_k = E[x_k|y_{-\infty}^{\infty}]$ and $e_k = x_k - \tilde{x}_k$. Making use of the hat-notation for the one-step predictor as in Lemma 5, i. e., $\hat{x}_k = E[x_k|x_{-\infty}^{k-1}]$ and $\hat{e}_k = E[e_k|e_{-\infty}^{k-1}]$, we find that

$$\begin{aligned} I(x(z);y(z)) &= I(x_k;\tilde{x}_k) + I(e_k; e_{-\infty}^{k-1}) - I(x_k; x_{-\infty}^{k-1}) && (27) \\ &= I(x_k;\tilde{x}_k) + I(e_k;\hat{e}_k) - I(x_k;\hat{x}_k). && (28) \end{aligned}$$

Thus, the average mutual information of the Gaussian ISI channel has been expressed in terms of three pairs of scalar-valued Gaussian variables. Before considering these pairs in more detail, we apply Lemmas 3 and 5 to (28) to get that

$$\begin{aligned} I(x(z);y(z)) &= \log\left(\frac{E[|x_k|^2]}{E[|x_k - \tilde{x}_k|^2]}\right) - \log\left(\frac{E[|x_k|^2]}{E[|x_k - \hat{x}_k|^2]}\right) + \log\left(\frac{E[|e_k|^2]}{E[|e_k - \hat{e}_k|^2]}\right) \\ &= \log\left(\frac{E[|x_k|^2]}{E[|e_k|^2]} \frac{E[|(i_x)_k|^2]}{E[|x_k|^2]} \frac{E[|e_k|^2]}{E[|(i_e)_k|^2]}\right) \\ &= \log\left(\frac{g_x}{g_e}\right). && (29) \end{aligned}$$

We can explicitly calculate $S_e(z)$ to be

$$\begin{aligned} S_e(z) &= S_x(z) - S_{xy}(z)S_y^{-1}(z)S_{yx}(z) && (30) \\ &= \frac{S_x(z)S_n(z)}{S_y(z)}, && (31) \end{aligned}$$

so that $g_e = g_x g_n / g_y$. Finally then, the result given in (9) is derived since $I(x(z);y(z))$ equals $\log(g_y/g_n)$.

We now look more closely at the three mutual-information terms that compose the right-hand side of (28). The first term is $I(x_k;\tilde{x}_k)$. Since $\tilde{x}(z)$ is the orthogonal projection of $x(z)$ onto $y(z)$, it can be expressed as $\tilde{x}(z) = S_{xy}(z)S_y^{-1}(z)y(z)$. Incorporating (5) and (6), we have that

$$\tilde{x}(z) = c(z)(a(z)x(z) + n(z)), \quad (32)$$

where

$$c(z) = \frac{S_x(z)a^*(1/z^*)}{a(z)S_x(z)a^*(1/z^*) + S_n(z)}. \quad (33)$$

We must be very careful, however, in our interpretation of this representation. Without further clarification, the channel in (32) represents $I(x(z);\tilde{x}(z))$, not $I(x_k;\tilde{x}_k)$. These are two very distinct channels. The first is a sequence-wise interpretation of the channel, while the second is a symbol-wise interpretation of the channel. Clearly, only this second interpretation is relevant to our discussion.

The second term in (28) is $I(e_k; \hat{e}_k)$. By Lemma 5, this is representable as

$$\hat{e}(z) = (1 - \phi_e^{-1}(z))e(z), \tag{34}$$

where $\phi_e^{-1}(z)$ is the causal, monic, minimum-phase whitening filter for $S_e(z)/g_e$, and of course $e(z) = x(z) - \tilde{x}(z)$. Again, we are interested in only the k^{th} input and the k^{th} output. And similarly, the third channel $I(x_k; \hat{x}_k)$ comes from the following

$$\hat{x}(z) = (1 - \phi_x^{-1}(z))x(z). \tag{35}$$

It turns out that the inputs and outputs of the channels in (32), (34), and (35) are inter-related in such a manner that allows them to be combined to form a channel that is equivalent as far as mutual information is concerned. The first and second channels are combined as shown in parts (a) and (b) of Figure 1. Since we are adding the outputs of the two channels, we must still verify that the mutual information of the overall channel is equal to the sum of the mutual informations of the two component channels.

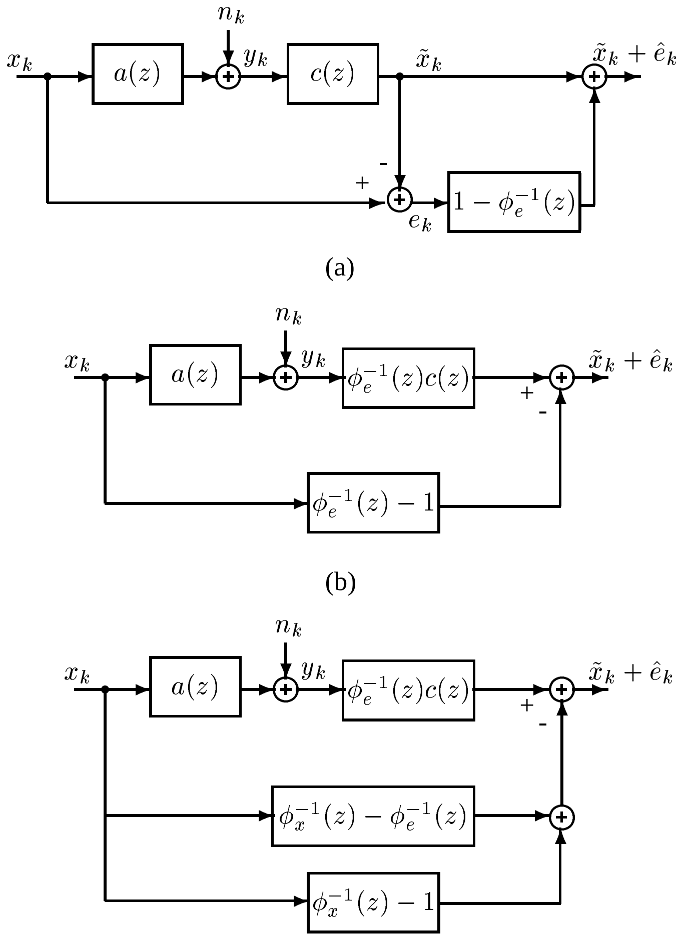
Lemma 6: When the channels $I(x_k; \tilde{x}_k)$ and $I(e_k; \hat{e}_k)$ are combined as shown in Figure 1, we have that

$$I(x_k; \hat{e}_k + \tilde{x}_k) = I(x_k; \tilde{x}_k) + I(e_k; \hat{e}_k). \tag{36}$$

This lemma can be proved directly. For example, see [10], [11], or [1, chap. 10] where the left-hand side is shown to equal $I(x(z); y(z))$ when $\{x_k\}$ is white. The result can also be shown to follow from a much more general statement concerning certain relationships of orthogonal-projection channels. The latter approach sidesteps the algebraic details associated with the direct proof. It also allows one to easily generalize the result to other Gaussian channels such as the multivariate ISI channel [15], the finite-dimensional ISI channel [16], and the synchronous Gaussian multiple-access channel [17].

It should be pointed out that channel shown in part (b) of Figure 1 is sometimes given as the perfect-feedback MMSE DFE structure for the Gaussian ISI channel [1, Sec. 10.2]. We see, however, that unless $\{x_k\}$ is a white process so that $I(x_k; \hat{x}_k) = 0$, this structure does not model the Gaussian ISI channel accurately. Equation (23) shows that it models the channel $I(x_k; y_{-\infty}^{\infty}, x_{-\infty}^{k-1})$ instead of the ISI channel of interest, $I(x_k; y_{-\infty}^{\infty} | x_{-\infty}^{k-1})$. To incorporate the third channel $I(x_k; \hat{x}_k)$, we can express $x(z)(1 - \phi_e^{-1}(z))$ as $x(z)(1 - \phi_x^{-1}(z)) + x(z)(\phi_x^{-1}(z) + \phi_e^{-1}(z))$. When this representation is incorporated into part (b) of Figure 1, the result is part (c) of the same figure. Note that the last parallel branch of this figure is is simply the third channel of interest. Thus, the subtraction of the third mutual-information term in (28) suggests the removal of this parallel branch. We know that

$$\phi_x^{-1}(z) - \phi_e^{-1}(z) = \phi_x^{-1}(z)\left(1 - \frac{\phi_y(z)}{\phi_n(z)}\right). \tag{37}$$



(a)

(b)

(c)

Fig. 1. Three equivalent forms of the channel $I(x_k; \tilde{x}_k + \hat{e}_k)$. Whenever $\{x_k\}$ is not white this is the same channel as $I(x_k; y_{-\infty}^{\infty}, x_{-\infty}^{k-1})$, but if $\{x_k\}$ is white then it is the same as the Gaussian ISI channel, i. e., $I(x_k; y_{-\infty}^{\infty} | x_{-\infty}^{k-1}) = I(x(z); y(z))$.

so that the combination of all three terms in (28) corresponds to the model pictured in Figure 2. It is easily seen that $\hat{e}_k + \tilde{x}_k - x_k + (i_x)_k = (i_x)_k - (i_e)_k$, and it is not hard to show that $I((i_x)_k; (i_x)_k - (i_e)_k) = I(x(z); y(z))$. In addition, we can show that $I(x_k; (i_x)_k - (i_e)_k) < I((i_x)_k; (i_x)_k - (i_e)_k)$ whenever $\{x_k\}$ is not white. Hence, Figure 2 only makes sense if we consider $(i_x)_k$ to be the input. Note that the symbol-wise interpretation of this channel is equivalent to the memoryless channel that is yielded by perfect interleaving. This, then, is the canonical MMSE DFE model, and it is the model that is typically assumed in the literature.[2]

Thus, we have shown that from the viewpoint of mutual information, there is no loss in generality in assuming that the receiver of the Gaussian ISI channel is the perfect-feedback MMSE DFE. Of course, perfect feedback means that the post-cursor ISI is completely removed

---

[2] The output $(i_x)_k - (i_e)_k$ is sometimes scaled by the constant α that yields the unbiased MMSE DFE, i. e., $E[\alpha((i_x)_k - (i_e)_k) | (i_x)_k] = (i_x)_k$ [10] [1, Sec. 10.2]. But as far as mutual information is concerned, there is no difference between the biased MMSE DFE and the unbiased MMSE DFE since $I((i_x)_k; \alpha[(i_x)_k - (i_e)_k]) = I((i_x)_k; (i_x)_k - (i_e)_k)$ for all non-zero α.

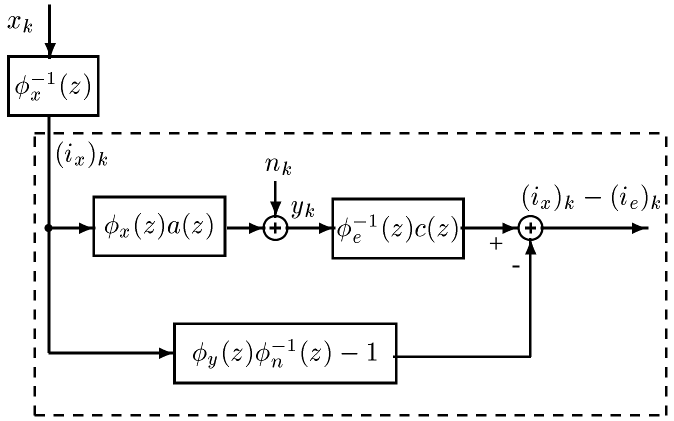


Fig. 2. The Gaussian ISI channel with a MMSE DFE receiver in the general case where $\{x_k\}$ is not white. The mutual information of the boxed channel $I((i_x)_k; (i_x)_k - (i_e)_k)$ is equivalent to that of the Gaussian ISI channel $I(x_k; y_{-\infty}^{\infty}|x_{\infty}^{k-1}) = I(x(z); y(z))$. Note that $c(z)$ and $S_e(z)$ are still given by (33) and (31), respectively.

by the receiver. This equivalence was derived by means of information-lossless orthogonal projections that convert the ISI channel, which is naturally viewed in a sequence-wise manner, to a channel that must be viewed symbol-wise. That is, the perfect-feedback MMSE DFE structure has built into it the fact that at time instance k, only those past symbols, i. e., $x_{-\infty}^{k-1}$, are available at the receiver. A sequence-wise interpretation violates this by inherently assuming knowledge of future symbols, i. e., x_{k+1}^{∞}, as well.

In [10, Prop. 1], the following direct part of the coding theorem for the perfect-feedback MMSE DFE is established.

Theorem 1 (Direct) Suppose we are given a Gaussian ISI channel with capacity C_{ISI}. If C_{DFE} denotes the capacity of the overall channel when the receiver is constrained to be the perfect-feedback MMSE DFE, then $C_{\mathrm{DFE}} \geq C_{\mathrm{ISI}}$. $\diamond$

We are able to provide the converse to this theorem. That is,

Theorem 2 (Converse) $C_{\mathrm{DFE}} \leq C_{\mathrm{ISI}}$. $\diamond$

This converse, was previously conjectured to be false because of the belief that perfect post-cursor ISI cancellation is in general an information-increasing operation [11] [8] [10]. This belief is based on the correctly observed fact that in general $I(i_x(z); i_x(z) - i_e(z)) > I((i_x)_k; (i_x)_k - (i_e)_k)$. But as we have indicated, these are two, very distinct channels. Our development has shown that the perfect-feedback MMSE DFE is a receiver structure that converts $I(x(z); y(z))$ into the latter. That is to say, the sequence of pairs $\{(i_x)_k; (i_x)_k - (i_e)_k\}$ when viewed symbol wise is sufficient for the pair $(x(z); y(z))$. The channel $I(i_x(z); i_x(z) - i_e(z))$ requires a sequence-wise interpretation of the MMSE DFE, an interpretation that is clearly incompatible with the MMSE DFE and, hence, also incompatible with the ISI channel. This highlights a potential danger of working with block diagrams of the Gaussian ISI channel with a MMSE DFE receiver, for such representations are not valid without the caveat that they *must* be viewed in symbol-wise manner.

IV. CONCLUDING REMARKS

We have seen fundamentally how it is that the MMSE DFE, under the assumption of perfect feedback, allows one to achieve the capacity of the Gaussian-noise channel with inter-symbol interference. The MMSE DFE effectively makes use of information-lossless projections to convert the ISI channel into a memoryless Gaussian channel. Our method of derivation can be applied to other channels (finite-dimensional ISI, multivariate ISI, and synchronous multiple-access) to yield the feedback equalizers given in [16], [15] as well as the multiuser decision-feedback receiver given in [17].

REFERENCES

[1] E. A. Lee and D. G. Messerschmitt, *Digital Communication*. Norwell, MA: Kluwer Academic Publishers, 1994.

[2] F. D. Neeser and J. L. Massey, “ Proper Complex Random Processes with Applications to Information Theory ,” *IEEE Trans. Inform. Theory*, vol. IT-39, pp. 1293–1302, Jul. 1993.

[3] J. G. Proakis, *Digital Communications*. New York, NY: McGraw-Hill Book Co., 1983.

[4] G. D. Forney Jr., “ Maximum-Likelihood Sequence Estimation of Digital Sequences in the Presence of Intersymbol Interference ,” *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 363–378, May 1972.

[5] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[6] W. Hirt and J. L. Massey, “ Capacity of the Discrete-Time Gaussian Channel with Intersymbol Interference ,” *IEEE Trans. Inform. Theory*, vol. IT-34, pp. 380–388, May 1988.

[7] S. Shamai (Shitz), L. H. Ozarow, and A. D. Wyner, “ Information Rates for Discrete-Time Gaussian Channel with Intersymbol Interference and Stationary Inputs ,” *IEEE Trans. Inform. Theory*, vol. IT-37, pp. 1527–1539, Nov. 1991.

[8] J. R. Barry, E. Lee, and D. G. Messerschmitt, “ Capacity Penalty Due to Ideal Zero-Forcing Decision-Feedback Equalization ,” *IEEE Trans. Inform. Theory*, vol. IT-42, pp. 1062–1071, Jul. 1996.

[9] J. M. Cioffi, G. P. Dudevoir, M. V. Eyuboglu, and J. Forney, G. D., “ MMSE Decision-Feedback Equalizers and Coding-Part I: Equalization Results ,” *IEEE Trans. Commun.*, vol. COM-43, pp. 2582–2594, Oct. 1995.

[10] J. M. Cioffi, G. P. Dudevoir, M. V. Eyuboglu, and G. D. Forney Jr., “ MMSE Decision-Feedback Equalizers and Coding-Part II: Coding Results ,” *IEEE Trans. Commun.*, vol. COM-43, pp. 2595–2604, Oct. 1995.

[11] S. Shamai (Shitz) and R. Laoria, “ The Intersymbol Interference Channel: Lower Bounds on Capacity and Channel Precoding Loss ,” *IEEE Trans. Inform. Theory*, vol. IT-42, pp. 1388–1404, Sep. 1996.

[12] M. S. Pinsker, *Information and Information Stability of Random Variables and Processes*. Holden-Day, Inc., 1964.

[13] I. M. Gelfand and A. M. Yaglom, “ Calculation of the Amount of Information About a Random Function Contained in another Such Function ,” *American Mathematical Society Translations*, vol. 12, pp. 199–246, 1959.

[14] A. Papoulis, *Probability, Random Variables, and Stochastic Processes, Second Edition*. New York, NY: McGraw-Hill Book Co., 1984.

[15] J. Yang and S. Roy, “ Joint Transmitter-Receiver Optimization for Multi-Input Multi-Output Systems with Decision Feedback ,” *IEEE Trans. Inform. Theory*, vol. IT-40, pp. 1334–1347, Sep. 1994.

[16] N. Al-Dhahir, J. M. Cioffi, and G. D. Forney Jr., “ A Canonical Finite-Length MMSE-DFE,” in *Proc. Thirty-third Allerton Conf. Commun., Contr. and Comput.,* Allerton, IL, pp. 598–607, Oct. 1995.

[17] M. K. Varanasi and T. Guess, “ Optimum Decision Feedback Multiuser Equalization with Successive Decoding Achieves the Total Capacity of the Gaussian Multiple-Access Channel ,” in *Proc. Thirty-First Asilomar Conf. Signals, Systems, and Computers* , Nov. 1997.